Lebiere, C., Biefeld, E., Archer, R., Archer, S., Allender, L., and Kelley, T. D. (2002).

Imprint/ACT-R: Integration of a Task Network Modeling Architecture with a Cognitive Architecture and its Application to Human Error Modeling.

In M. J. Chinni (Ed). 2002 Military, Government and Aerospace Simulation, 34(3), 13-19. San Diego: Society for Modeling and Simulation International.

IMPRINT/ACT-R: Integration of a Task Network Modeling Architecture with a Cognitive Architecture and its Application to Human Error Modeling

Christian Lebiere
Eric Biefeld
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
cl@cmu.edu
biefeld@cs.cmu.edu

Rick Archer
Sue Archer
Micro Analysis & Design
4900 Pearl East Circle
Boulder, CO 80301
rarcher@maad.com
sarcher@maad.com

Laurel Allender
Troy D. Kelley
Army Research Laboratory
Aberdeen Proving Grounds, MD 21005
lallende@arl.army.mil
tkelley@arl.army.mil

Keywords: Task network model, cognitive architecture, IMPRINT, ACT-R, error modeling

Abstract

This paper describes ongoing efforts to integrate IMPRINT (IMproved Performance Research INtegration Tool), a task network modeling architecture with ACT-R (Adaptive Character of Thought — Rational), a hybrid cognitive architecture. IMPRINT consists of a set of automated aids to conduct human performance analyses built on top of the Micro Saint task network modeling environment. ACT-R combines a goal-directed production system with a subsymbolic activation calculus that tunes itself to the structure of the environment using Bayesian learning mechanisms. Because ACT-R and IMPRINT were targeted at different behavioral levels, they perfectly complement each other. IMPRINT is focused on the task level, how high-level functions break down into smaller-scale tasks and the logic by which those tasks follow each other to accomplish those functions. ACT-R is targeted at the ÒatomicÓ level of thought, the individual cognitive, perceptual and motor acts that take place at the sub-second level. Goals in ACT-R correspond directly to tasks in IMPRINT, providing a natural integration level. Certain tasks in an IMPRINT task network can be implemented as ACT-R models, combining the cognitive accuracy of a cognitive architecture with the tractability and ease of design of task networks.

A hybrid IMPRINT/ACT-R model works as follows. The IMPRINT model specifies the network of tasks and includes the definition of how higher-order functions are decomposed into tasks and the logic by which these tasks are composed together. For certain tasks, IMPRINT sends to ACT-R over a Component Object Model (COM) link the state of variables providing a detailed description of that task. ACT-R then creates a goal corresponding to that task, with the components of the goal set to the description of the task. The ACT-R model for that goal is then run, producing detailed cognitive predictions including latency of the run, whether an error occurred, etc. Those results are then passed back over the same COM link to IMPRINT, which uses them as parameters of the task to advance the task network model. We describe an application of this hybrid modeling to the prediction of human errors that lead to runway incursions. Finally, we discuss future extensions of our work, including the use of a standardized High Level Architecture (HLA) link to handle communications between IMPRINT and ACT-R and the extension of the task parameters exchanged to include workload predictions.

INTRODUCTION

The primary benefit of this effort is in taking advantage of the strengths of each tool in order to increase the fidelity of human performance models, without unnecessarily burdening the model developer. Additionally, this project exploits the synergy between the computer science and cognitive science communities. This synergy will promote the advancement of human performance modeling approaches and tools through the selective application of artificial intelligence technology

IMPRINT

IMPRINT (IMproved Performance Research INtegration Tool) was developed for ARL HRED to conduct human performance analyses very early in the acquisition of a proposed weapon system. It consists of a set of automated aids to assist researchers in conducting human performance analyses.

IMPRINT assists a user in estimating the likely performance of a new system by facilitating the construction of flow models that describe the scenario, the environment, and the mission that must be accomplished. Since it is typically easier to describe the mission by breaking it into smaller ÒsubÓfunctions than trying to describe the mission as a whole, users build these models by breaking down the mission into a network of functions. Each of the functions is then further broken down into a network consisting of other functions and tasks. Then, a user estimates the time is will take to perform each task and the likelihood that it will be performed accurately.

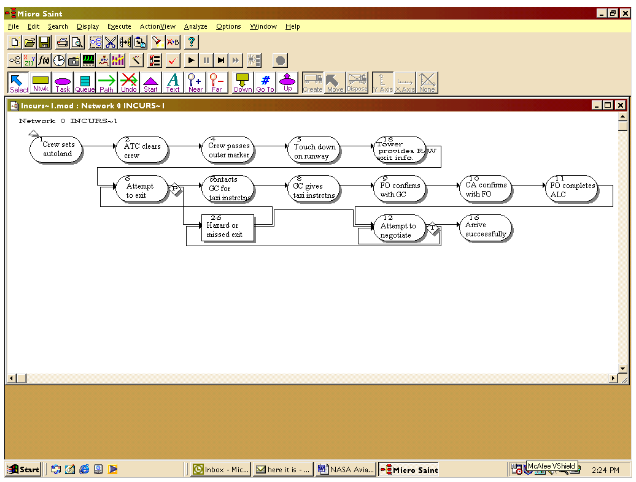

Figure 1. IMPRINT Task Network

Finally, by executing a simulation model of the mission multiple times, you can study the range of results that occur. A description of the variability of each element can be obtained for further analysis. Additionally, at the completion of the simulation, IMPRINT can compare the minimum acceptable mission performance time and accuracy to the predicted performance. This will determine whether the mission met its performance requirements.

IMPRINT has been used successfully to predict human performance in complex and dynamic operational environments. It has been shown to be easy to use, and fairly quick to apply. It does not include an embedded model of cognitive or psychological processes. Rather, it relies on the modeler to specify and implement these constructs. In our proposed approach for this effort, these constructs will be provided through a link to ACT-R.

IMPRINT is a government-owned product developed by MA&D, and whose progress has been directed and monitored by our team member, ARL HRED. IMPRINT was implemented in C on the Windows platform. It includes a graphical user interface for model authoring, a library of existing weapon systems, expandable function libraries, data collection and display modules, and built-in optimization and animation tools for simulation development. These capabilities provide an easy-to-use interface for the development of simulation models that can be used to study and assess human processes.

ACT-R

ACT-R [Anderson and Lebiere 1998] was developed at Carnegie Mellon University under sponsorship from the Office of Naval Research. ACT-R runs on MacOS, Windows and Unix. A number of user tools are available, including a graphical environment to author and run ACT-R models, an adaptive web tutorial for learning how to model in ACT-R, a parameter optimizer that automates the task of model fitting and a multi-model extension that enables multiple ACT-R models to run concurrently and communicate with each other and with an interactive simulation. ACT-R is open-source and all software, models and tools are freely available on the web at the ACT-R web site http://act.psy.cmu.edu.

ACT-R is a production system theory that tries to model the steps of cognition by a sequence of production rules that fire to coordinate retrieval of information from the environment and from memory. It is a cognitive architecture that can be used to model a wide range of human cognition. It has been used to model tasks as simple as memory retrieval and visual search to tasks as complex as learning physics and designing psychology experiments. In all domains, ACT-R is distinguished by the detail and fidelity with which it models human cognition. It predicts what happens cognitively every few hundred milliseconds in performance of a task. ACT-R is situated at a level of aggregation considerably above basic brain processes but considerably below significant tasks like air-traffic control. The newest version of ACT-R has been designed to be more relevant to tasks that are being performed under conditions of time pressure and high information-processing demand.

Figure 2 displays the information flow in the ACT-R system. There are essentially three memories -- a goal stack that encodes the hierarchy of intentions guiding behavior, a procedural memory containing production rules, and a declarative memory containing chunks. Productions are condition-action pairs that determine which basic cognitive actions can be taken and when. Chunks are knowledge structures holding a small set of elements (e.g. 3+4=7) in labeled slots.

Access to these memories is coordinated around the current goal that represents the focus of attention. The current goal can be temporarily suspended when a new goal is pushed on the stack. The current goal can be popped in which case the next goal will be retrieved from the stack. Productions are selected to fire through a conflict resolution process that chooses one production from among the productions that match the current goal. The selected production can cause actions to be taken in the outside world, can transform the current goal (possibly resulting in pushes and pops to the stack), and can make retrieval requests of declarative memory (e.g., Òwhat is the sum of 3 and 4?Ó). The retrieval result (e.g., Ò7Ó) can be returned to the goal. The arrows in Figure 2 also describe how new declarative chunks and productions are acquired. Chunks can be added to declarative memory either as popped goals reflecting the solutions to past problems or as perceptions from the environment. Productions are created from declarative chunks through a process called production compilation which takes an encoding of an execution trace resulting from multiple production firings and produces a new production that implements a generalization of that transformation in a single production cycle.

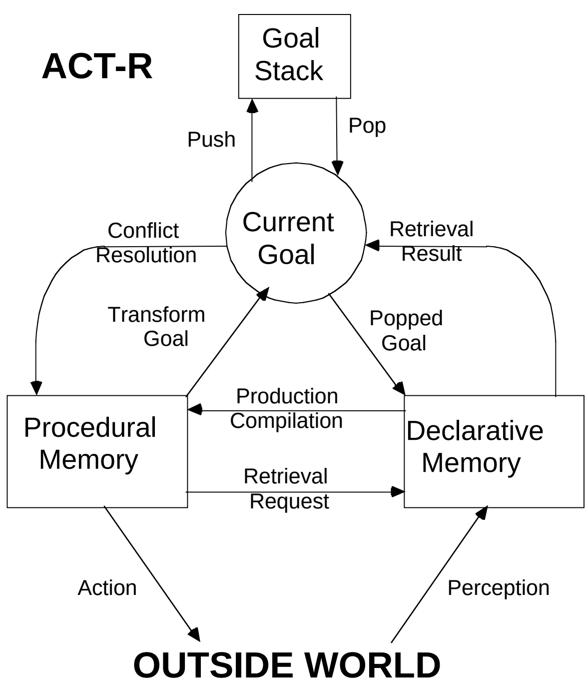

Figure 2. The overall flow of control in ACT-R.

ACT-R also has a subsymbolic level in which continuously varying quantities are processed, often in parallel, to produce much of the qualitative structure of human cognition. These subsymbolic quantities participate in neural-like activation processes that determine the speed and success of access to chunks in declarative memory as well as the conflict resolution among production rules. ACT-R also has a set of learning processes that can modify these subsymbolic quantities.

The activation of a declarative memory chunk determines its availability. The context activation is a function of the attentional weight given to the current goal. The base level activation of a chunk is learned by an architectural mechanism according to Bayesian statistics to reflect the past history of use of the information contained in the chunk. This equation produces the Power Law of Forgetting as well as the Power Law of Learning.

When trying to retrieve a chunk to instantiate a production, ACT-R selects the chunk with the highest activation. That activation includes a random component that provides stochasticity to memory retrieval and hence to the modelÕs behavior, as well as a similarity-based matching component, which provides generalization and robustness. Thus, ACT-R is capable both of errors of omission, in which a chunk cannot be retrieved because its activation cannot reach a threshold, and errors of commission, in which the wrong chunk is retrieved instead of the correct one [Lebiere et al., 1994]. The retrieval time of a chunk is an exponential function of its activation, providing a fine-grained account of the time scale of memory access. The total time of selecting and applying a production is determined by executing the actions of a productionÕs action part, whereby a value of 50 ms is typically assumed for elementary internal actions. External actions, such as pressing a key, usually have a longer latency determined by the ACT-R/PM perceptual-motor modules. In summary, subsymbolic activation processes in ACT-R make a chunk active to the degree that past experience and the present context (as given by the current goal) indicates usefulness at this particular moment.

TASK

As a practical application of the IMPRINT and ACT-R integration, a complex and dynamic task was selected for a modeling effort. Researchers with the National Aeronautics and Space Administration (NASA) were interested in developing models of pilot navigation while taxiing from a runway to a gate. Research on pilot surface operations had shown that pilots can commit numerous errors during taxi procedures [Hooey and Foyle 2001]. NASA was hoping to reduce the number and scope of pilot error during surface operations by using information displays that would improve the pilots' overall situation awareness.

NASA researchers provided the IMPRINT and ACT-R modeling teams with data describing pilot procedures during pre-landing and surface taxi operations. This data included videotapes of pilots in the NASA Ames Advanced Concept Flight Simulator (ACFS) which is a simulated cockpit capable of duplicating pilot taxiing operations. A detailed, scaled, map of Chicago's O'Hare airport was also provided which included runway signage. Other types of documentation was provided to give the IMPRINT and ACT-R modeling team the information necessary to duplicate runway taxiing behavior by pilots.

The IMPRINT and ACT-R modeling teams used the scaled map of Chicago's O'Hare airport to estimate the time between runway taxi turns. IMPRINT handled the higher level, task oriented parts of the taxiing and landing operations, (i.e. turning, talking on radio, looking at instrumentation) while ACT-R handled the more cognitive and decision making parts of the task (i.e. remembering where to turn, remembering the taxi route). By using the scaled map of the airport, the IMPRINT and ACT-R teams were able to determine the amount time between each taxi turn (based on an estimated plane speed which was correlated with the simulated speeds from the video tape data) and then use that data to estimate the decay rate for the list of memory elements (i.e. runway names) that the pilot would have to remember.

Using this integrated architecture allowed us to be able to represent a complex, dynamic task and by using the strengths of each architecture, the modeling process was enhanced and streamlined.

INTEGRATION

Because ACT-R and IMPRINT were targeted at different behavioral levels, they perfectly complement each other. IMPRINT is focused on the task level, how high-level functions break down into smaller-scale tasks and the logic by which those tasks follow each other to accomplish those functions. ACT-R is targeted at the ÒatomicÓ level of thought, the individual cognitive, perceptual and motor acts that take place at the sub-second level. As seen in Figure 2, ACT-R is centered on the concept of the current goal. At each cycle, a production will be chosen that best applies to the goal, knowledge might be retrieved from declarative memory and perceptual and motor actions taken. Those cycles will repeat until the current goal is solved, at which point it is popped and another one is selected. The ACT-R theory specifies in detail the performance and learning that takes place at each cycle within a specific goal, but has comparatively little to say about the selection of those goals. Since goals in ACT-R closely correspond to tasks in IMPRINT, that weakness matches perfectly IMPRINTÕs strength. Conversely, since IMPRINT requires the characteristics of each task to be specified as part of the model, ACT-R can be used to generate those detailed characteristics in a psychologically plausible way without requiring extensive data collection.

An IMPRINT model specifies the network of tasks used to accomplish the functions targeted by the model, e.g. landing a plane and taxiing safely to the gate. The network specifies how higher-order functions are decomposed into tasks and the logic by which these tasks are composed together. As input, it takes the distribution of times to complete the task and the accuracy with which the task is completed. It can also take as input the workload generated by each task. Additional inputs include events generated by the simulation environment. Finally, a number of additional general parameters such as personnel characteristics, level of training and familiarity and environmental stressors can be specified. IMPRINT specifies the performance function by which these parameters modulate human performance. The outputs include mission performance data such as time and accuracy as well as aggregate workload data.

An ACT-R model specifies the knowledge structures such as declarative chunks and production rules that constitute the user knowledge relevant to the tasks targeted by the model. It also specifies the goal structures reflecting the task structure and the architectural and prior knowledge parameters that modulate the modelÕs performance. For each goal on which ACT-R is focused (i.e. made the current goal), it generates a series of sub-second cognitive, perceptual and motor actions. The result of those actions is the total time to accomplish the goal, as well as how the goal was accomplished, including any error that might result. Errors in ACT-R originate from a broad range of sources. They include memory failures, including the failure to retrieve a needed piece of information or the retrieval of the wrong piece of information, choice failures, including the selection of the wrong production rule, and attentional failures, such as the failure to detect the salient piece of information by the perceptual modules. While those errors could arise because of faulty symbolic knowledge (either declarative or procedural), it is often not the case, especially in domains that involve highly trained crews. More often, those errors occur because the subsymbolic parameters associated with chunks or productions does not allow the model to access them reliably or quickly enough to be deployed in the proper situation. Moreover, because those parameters vary stochastically and their effect is amplified by the interaction with a dynamic environment, those times and errors will not be deterministic but will vary with each execution, as is the case for human operators. Thus the ACT-R model for a particular goal can be run whenever IMPRINT selects the corresponding task to generate the time and error distribution for that task in a manner that reflects the myriad cognitive, perceptual and motor factors that enter into the actual performance of the task. ACT-R can also generate workload estimates for each task that reflect the cognitive demands of the actions taken to perform the task [Lebiere et al., 2001; Lebiere, 2001].

MODEL

IMPRINT

The IMPRINT task network model consists of the tasks that the Captain and the First Officer perform from the time that the airplane approaches the airport from about 12 miles out until the airplane either commits a taxi navigation error or arrives at the correct terminal gate without committing an error. The tasks in the model are grouped into three general segments. Prior to the execution of the first segment, an initialization task communicates all of the correct information about where the airplane will be directed to land and the taxi route information to the gate destination.

The first segment is the approach, which begins with the tasks that the Captain and First officer must perform in preparation for landing. It is during this approach segment that Air Traffic Control communicates runway landing information to the crew.

Next is the Òroll outÓ segment in which the crew lands the aircraft and proceeds down the runway until the correct runway turn-off has been reached. It is during the runway roll out segment that the taxi route and gate information is communicated to the crew by the control tower. As the crewmembers observe their displays and communicate with one another, this information is shared with the ACT-R model via shared variables through COM. As the aircraft approaches each potential runway turn-off, information about whether runway signage has been noticed or communicated is also passed to the ACT-R model. In turn, ACT-R passes information back to IMPRINT about whether to make a turn and in which direction to turn.

After turning off the runway the Òtaxi inÓ segment takes the aircraft to gate. Information about taxiway signage, the crewÕs use of displays and their communication is passed to the ACT-R model. ACT-R again passes information back to IMPRINT as to whether a turn should be made, whether the aircraft should stop and wait or whether it should proceed to the next taxiway intersection.

Aside from the normal procedures that are being performed by the crew, other events such as radio communications or taxiway traffic can occur. If the aircraft makes an error in which turn to make or in which direction to turn, the simulation is terminated. The model can be executed many times to predict the likelihood of navigation errors.

ACT-R

In the spirit of concentrating on the areas where cognitive accuracy is most critical, the ACT-R model focused on the task of memorizing and recalling the list of taxiways to follow after landing. This task is similar to the cognitive psychology task of list learning, for which an ACT-R model had already been developed [Anderson et al., 1998]. We adapted that model to the task at hand while preserving its fundamental representation and parameters, thus eliminating degrees of freedom and inheriting that modelÕs empirical validation. The taxiways turns were represented by two chunks each, one indicating the name of the taxiway and the other holding the direction to turn. The ACT-R model was called for the initial memorization of the list of taxiways and then each time the aircraft approached a taxiway intersection.

Results

The model could reproduce the full range of errors observed in human pilots. Omission errors occurred when a chunk holding a turn could not be recalled because of time-based decay or activation noise. The resulting error would be a missed turn. Two kinds of commission errors could occur. The first kind would result in the wrong chunk recalled because of interference, similarity-based partial matching, priming or activation noise. This would cause the model to schedule a turn on the wrong taxiway. The second kind of commission error would happen when the wrong direction chunk was retrieved, again for reasons of interference or noise. This would result in a turn on the correct taxiway but in the wrong direction. A small sample of 5 model runs produced a diversity of outcomes. Two runs produced all the correct turns; two runs ended with a missed turn and one run ended with a turn in the wrong direction. While longer sample of runs remain to be collected and analyzed, this indicates that our model can capture the range of real-world outcomes.

DISCUSSION

Many issues remain to achieve a successful seamless integration of the two products. One significant issue is performance. In order for the two tools to be effective, the project team must find ways to enable the products to work in parallel, while sharing many variables (e.g., clock time) Òon the fly.Ó Since each product has its own idea of what clock time is, and how it should be used, this represents a substantial challenge. Another relevant issue is the concept of workload. IMPRINT contains algorithms that are used to represent the level of effort needed to perform tasks as they occur in the network. ACT-R could potentially use these measures as a way of ÒmoderatingÓ the attributes of the memory model. The project team is currently working on these and other important challenges under a variety of sponsors, including the Air Force Research Laboratory, the Army Research Laboratory and NASA.

CONCLUSION

Our project team is very encouraged by the progress achieved in the project we describe in this paper. The integrated model appears to increase our ability to predict human error, and appears to be a parsimonious solution to the problem of selecting an appropriate level of fidelity for human performance models. However, there is still much to do before a seamless integration is achieved.

REFERENCES

Anderson, J. R. and C. Lebiere. 1998. *The atomic components of thought.* Erlbaum, Mahwah, NJ.

Anderson, J. R.; D. Bothell; C. Lebiere; M. Matessa. 1998. ÒAn integrated theory of list memory.Ó *Journal of Memory and Language, 38,* 341-380.

Hooey, B. L and D. C. Foyle. 2001. ÒA post-hoc analysis of navigation errors during surface operations. Identification of contributing factors and mitigating strategies.Ó In *Proceedings of the 11th Symposium on Aviation Psychology*. Columbus, Ohio: Ohio State University.

Lebiere, C.; Anderson, J. R.; Reder, L. M. 1994. ÒError modeling in the ACT-R production system.Ó In *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ, 555-559

Lebiere, C.; Anderson, J.R.; Bothell, D. 2001. ÒMulti-tasking and cognitive workload in an ACT-R model of a simplified air traffic control task.Ó In *Proceedings of the Tenth Conference on Computer Generated Forces and Behavior Representation*. Norfolk, VA.

Lebiere, C. 2001. ÒA theory-based model of cognitive workload and its applications.Ó In *Proceedings of the 2001 Interservice/Industry Training, Simulation and Education Conference*. Orlando, FL.

BIOGRAPHY

CHRISTIAN LEBIERE is a Research Scientist in the Human-Computer Interaction Institute at Carnegie Mellon University. During his graduate career, he worked on the development of connectionist models, including the Cascade-Correlation neural network learning algorithm that has been used in hundreds of scientific, technical and commercial applications. Since 1990, he has worked on the development of the ACT-R hybrid cognitive architecture and is co-author with John R. Anderson of the 1998 book *The Atomic Components of Thought*. His main research interest is cognitive architectures and their applications to psychology, artificial intelligence, human-computer interaction, decision-making, game theory, and computer-generated forces.

ERIC BIEFELD is a research programmer at Carnegie Mellon University.

RICK ARCHER is a Principal Analyst at Micro Analysis and Design, Inc. located in Boulder, CO. Mr. Archer has been involved in simulation and modeling of human performance for seventeen years. He is currently the project leader for the integration of IMPRINT and ACT-R.

SUE ARCHER is the Director of Operations at Micro Analysis and Design. She is a member of Modeling and Simulation International, and is the responsible for the development of the IMPRINT product.

LAUREL ALLENDER is a Research Psychologist at the Human Research and Engineering Directorate of the Army Research Laboratory at Aberdeen Proving Ground, MD. Her research is concentrated on the modeling and analysis tool IMPRINT (Improved Performance Research Integration Tool) which has been transitioned to over 120 sites, applied to significant Army programs, and is related to similar Air Force, Navy, and foreign R&D efforts. Dr. Allender received a Ph.D. in cognitive psychology from Rice University, Houston, TX in 1987.

TROY KELLEY is an Engineering Psychologist at the Human Research and Engineering Directorate (HRED) of the Army Research Laboratory (ARL) at Aberdeen Proving Ground, MD. His recent work has concentrated on the understanding of human cognition, especially in areas of problem solving, strategy selection, and decision-making. Currently he is heading a program called MINDSS (Modeling and Integration of Neurological Dynamics with Symbolic Structures) which the goal of modeling human cognitive processes on ARL’s HPC (High Performance Computing) assets in cooperation with major U.S. universities.